Inferring parts of speech for lexical mappings via the Cyc KB

Tom O'Hara[‡], **Stefano Bertolo, Michael Witbrock, Bjørn Aldag, Jon Curtis**, with **Kathy Panton**, **Dave Schneider**, and **Nancy Salay**

[‡]Computer Science Department
New Mexico State University
Las Cruces, NM 88001
tomohara@cs.nmsu.edu

Cycorp, Inc.
Austin, TX 78731
{bertolo,witbrock,aldag}@cyc.com
{jonc,panton,daves,nancy}@cyc.com

Abstract

We present an automatic approach to learning criteria for classifying the parts-of-speech used in lexical mappings. This will further automate our knowledge acquisition system for non-technical users. The criteria for the speech parts are based on the types of the denoted terms along with morphological and corpus-based clues. Associations among these and the parts-of-speech are learned using the lexical mappings contained in the Cyc knowledge base as training data. With over 30 speech parts to choose from, the classifier achieves good results (77.8% correct). Accurate results (93.0%) are achieved in the special case of the mass-count distinction for nouns. Comparable results are also obtained using OpenCyc (73.1% general and 88.4% mass-count).

1 Introduction

In semantic lexicons, the term *lexical mapping* describes the relation between a concept and a phrase used to refer to it (Onyshkevych and Nirenburg, 1995; Burns and Davis, 1999). Lexical mappings include associated syntactic information, in particular, part of speech information for phrase headwords. The term *lexicalize* will refer to the process of producing these mappings, which are referred to as *lexicalizations*. Selecting the part of speech for the lexical mapping is required so that proper inflectional variations can be recognized and generated for the term. Although producing lexicalizations is often a straightforward task, there are many cases that can pose problems, especially when fine-grained speech part categories are used.

For example, the headword 'painting' is a verb in the phrase "painting for money" but a noun in the phrase "painting for sale." In cases like this, semantic or pragmatic criteria, as opposed to syntactic criteria only, are necessary for determining the proper part of speech. The headword part of speech is important for correctly identifying phrasal variations. For instance, the first term can also occur in the same sense in "paint for money." However, this does not hold for the second case, since "paint for sale" has an entirely different sense (i.e., a substance rather than an artifact).

When lexical mappings are produced by naive users, such as in DARPA's Rapid Knowledge Formation (RKF) project, it is desirable that technical details such as the headword part of speech be inferred for the user. Otherwise, often complex and time-consuming clarification dialogs might be necessary in order to rule out various possibilities. For example, Cycorp's Dictionary Assistant was developed for RKF in order to allow non-technical users to specify lexical mappings from terms into the Cyc knowledge base (KB). Currently, when a new type of activity is described, the user is asked a series of questions about the ways of referring to the activity. If the user enters the phrase "painting for money," the system asks whether the phrases "paint for money" and "painted for money" are suitable variations in order to determine whether 'painting' should be treated as a verb. Users find such clarification dialogs distracting, since they are more interested in entering domain rather than linguistic knowledge. Regardless, it is often very difficult to produce prompts that make the distinction intelligible to a linguistically naive user.

A special case of the lexicalization speech part classification is the handling of the mass-count distinction. Having the ability to determine if a concept takes a mass or count noun is useful not only for parsing, but also for generation of grammatical English. For example, automatically gener-

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Inferring parts of speech for lexical mappings via the Cyc KB	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Computer Science,New Mexico State University,PO Box 30001,Las Cruces,NM,88003-8001**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**7**	

Collection: **PhysicalDevice**
Microtheory: ArtifactGVocabularyMt
isa: ExistingObjectType
genls: Artifact ComplexPhysicalObject
SolidTangibleProduct
Microtheory: ProductGMt
isa: ProductType

Figure 1: **Type definition for *PhysicalDevice*.**

ated web pages (e.g., based on search terms) occasionally produce ungrammatical term variations because this distinction is not addressed properly.

Although learner dictionaries provide information on the mass-count distinction, they are not suitable for this task because different senses of a word are often conflated in the definitions for the sake of simplicity. In cases like this, the word or sense might be annotated as being both count and mass, perhaps with examples illustrating the different usages. This is the case for 'chicken' from the Cambridge International Dictionary of English (Procter, 1995), defined as follows:

> a type of bird kept on a farm for its eggs or its meat, or the meat of this bird which is cooked and eaten

This work describes an approach for automatically inferring the parts of speech for lexical mappings, using the existing lexical assertions in the Cyc KB. We are specifically concerned with selecting parts of speech for entries in a semantic lexicon, not about determining parts of speech in context. After an overview of the Cyc KB in the next section, Section 3 discusses the approach taken to inferring the part of speech for lexicalizations. Section 4 then covers the classification results. This is followed by a comparison to related work in Section 5.

2 Cyc knowledge base

In development since 1984, the Cyc knowledge base (Lenat, 1995) is the world's largest formalized representation of commonsense knowledge, containing over 120,000 concepts and more than a million axioms.[1] Cyc's upper ontology describes the most general and fundamental of distinctions (e.g., tangibility versus intangibility). The lower ontology contains facts useful for particular applications, such as web searching, but not necessarily required for commonsense reasoning (e.g., that "Dubya" refers to *President George W. Bush*). The KB also includes a broad-coverage English lexicon mapping words and phrases to terms throughout the KB. A subset of the Cyc KB including parts of the English lexicon has been made freely available as part of OpenCyc (www.opencyc.org).

[1]These figures and the results discussed later are based on Cyc KB version 576 and OpenCyc KB version 567.

2.1 Ontology

Central to the Cyc ontology is the concept *collection*, which corresponds to the familiar notion of a set, but with membership intensionally defined (so distinct collections can have identical members, which is impossible for sets). Every object in the Cyc ontology is a member (or *instance*, in Cyc parlance) of one or more collections. Collection membership is expressed using the predicate (i.e., relation-type) *isa*, whereas collection subsumption is expressed using the transitive predicate *genls* (i.e., generalization). These predicates correspond to the set-theoretic notions *element of* and *subset of* respectively and thus are used to form a partially ordered hierarchy of concepts. For the purposes of this discussion, the *isa* and *genls* assertions on a Cyc term constitute its *type definition*.

Figure 1 shows the type definition for *PhysicalDevice*, a prototypical denotatum term for count nouns. The type definition of *PhysicalDevice* indicates that it is a collection that is a specialization of *Artifact*, etc. As is typical for terms referred to by count nouns, it is an instance of the collection *ExistingObjectType*.

Figure 2 shows the type definition for *Water*, a prototypical denotation for mass nouns. Although the *asserted* type information for *Water* does not convey any properties that would suggest a mass noun lexicalization, the *genls* hierarchy of collections does. In particular, the collection *ChemicalCompoundTypeByChemicalSpecies* is known to be a specialization of the collection *ExistingStuffType*, via the transitive properties of *genls*. Thus, by virtue of being an instance of *ChemicalCompoundTypeByChemicalSpecies*, *Water* is known to be an instance of *ExistingStuffType*. This illustrates that the decision procedure for the lexical mapping speech parts needs to consider not only *asserted*, but also *inherited* collection membership.

2.2 English lexicon

Natural language lexicons are integrated directly into the Cyc KB (Burns and Davis, 1999). Though several lexicons are included in the KB, the English lexicon is the only one with general coverage. The mapping from nouns to concepts is done using one of two general strategies, depending on whether the

Collection: **Water**
Microtheory: UniversalVocabularyMt
isa: ChemicalCompoundTypeByChemicalSpecies
Microtheory: UniversalVocabularyMt
genls: Individual
Microtheory: NaivePhysicsVocabularyMt
genls: Oxide

Figure 2: **Type definition for *Water*.**

mapping is from a name or a common noun phrase. Several different binary predicates indicate name-to-term mappings, with the name represented as a string. For example,

(nameString HEBCompany "HEB")

A *denotational assertion* maps a phrase into a concept, usually a collection. The phrase is specified via a lexical word unit (i.e., lexeme concept) with optional string modifiers. The part of speech is specified via one of Cyc's *SpeechPart* constants. Syntactic information, such as the wordform variants and their speech parts, is stored with the Cyc constant for the word unit. For example, *Device-TheWord*, the Cyc constant for the word 'device,' has a single syntactic mapping since the plural form is inferable:

Constant: Device-TheWord
Microtheory: GeneralEnglishMt
isa: EnglishWord
posForms: CountNoun
singular: "device"

The simplest type of denotational mapping associates a particular sense of a word with a concept via the *denotation* predicate. For example,

(denotation Device-Word CountNoun 0 PhysicalDevice)

This indicates that sense 0 of the count noun 'device' refers to *PhysicalDevice* via the associated wordforms "device" and "devices."

To account for phrasal mappings, three additional predicates are used, depending on the location of the headword in the phrase. These are *compoundString*, *headMedialString*, and *multiWordString* for phrases with the headword at the beginning, the middle, and the end, respectively. For example,

(compoundString Buy-TheWord ("down") Verb BuyDown)

	Usage	
Predicate	OpenCyc	Cyc
multiWordString	1123	24606
denotation	2080	16725
compoundString	318	2226
headMedialString	200	942
total	3721	44499

Table 1: **Denotational predicate usage in Cyc English lexicon.** This excludes slang and jargon.

	Usage	
SpeechPart	OpenCyc	Cyc
CountNoun	2041	21820
MassNoun	566	9993
Adjective	262	6460
Verb	659	2860
AgentiveNoun	81	1389
ProperCountNoun	16	906
Adverb	50	310
ProperMassNoun	1	286
GerundiveNoun	7	275
other	39	185
total	3721	44499

Table 2: **Most common speech parts in denotational assertions.** The *other* entry covers 20 infrequently used cases.

This states that "buy down" refers to *BuyDown*, as do "buys down," "buying down," and "bought down" based on the inflections of the verb 'buy.'

Table 1 shows the frequency of the various predicates used in the denotational assertions, excluding lexicalizations that involve technical, informal or slang terms. Table 2 shows the most frequent speech parts from these assertions. This shows that nearly 50% of the cases use *CountNoun* for the headword speech part and that about 25% use *MassNoun*. This subset of the denotational assertions forms the basis of the training data used in the mass versus count noun classifier, as discussed later. Twenty other speech parts used in the lexicon are not shown. Several of these are quite specialized (e.g., *QuantifyingIndexical*) and not very common, mainly occurring in fixed phrases. The full speech part classifier handles all categories.

3 Inference of default part of speech

Our method of inferring the part of speech for lexicalizations is to apply machine learning techniques over the lexical mappings from English words or

phrases to Cyc terms. For each target denotatum term, the corresponding types and generalizations are extracted from the ontology. This includes terms for which the denotatum term is an instance or specialization, either explicitly asserted or inferable via transitivity. For simplicity, these are referred to as *ancestor terms*. The association between the lexicalization parts of speech and the common ancestor terms forms the basis for the main criteria used in the lexicalization speech part classifier and the special case for the mass-count classifier. In addition, this is augmented with features indicating whether known suffixes occur in the headword as well as with corpus statistics.

3.1 Cyc ancestor term features

There are several possibilities in mapping the Cyc ancestor terms into a feature vector for use in machine learning algorithms. The most direct method is to have a binary feature for each possible ancestor term, but this would require about ten thousand features. To prune the list of potential features, frequency considerations can be applied, such as taking the most frequent terms that occur in type definition assertions. Alternatively, the training data can be analyzed to see which reference terms are most correlated with the classifications.

For simplicity, the frequency approach is used here. The most-frequent 1024 atomic terms are selected, excluding terms used for bookkeeping purposes (e.g., *PublicConstant*, which mark terms for public releases of the KB); half of these terms are taken from the *isa* assertions, and the other half from the *genls* assertions. These are referred to as the *reference terms*. For instance, *ObjectType* is a type for 21,108 of the denotation terms (out of 44,449 cases), compared to 20,283 for *StuffType*. These occur at ranks 13 and 14, so they are both included. In contrast, *SeparationEvent* occurs only 185 times as a generalization term at rank 522, so it is pruned. See (O'Hara et al., 2003) for more details on extracting the reference term features.

3.2 Morphology and corpus-based features

In English, the suffix for a word can provide a good clue as to the speech part of a word. For example, agentive nouns commonly end in '-or' or '-er.' Features to account for this are derived by seeing whether the headword ends in one of a predefined set of suffixes and adding the suffix as a value to an enumerated feature variable corresponding to suffixes of the given length. Currently, the suffixes used are the most-common two to four letter sequences found in the headwords.

Feature	Search Pattern
singular	⟨*singular*⟩
plural	⟨*plural*⟩
count	"many ⟨*plural*⟩" or "several ⟨*plural*⟩"
mass	"much ⟨*singular*⟩" or "several ⟨*singular*⟩"
verb	"must ⟨*head*⟩" or "could ⟨*head*⟩"
adverb	"did ⟨*head*⟩" or "do ⟨*head*⟩" or "does ⟨*head*⟩" or "so ⟨*head*⟩" or "has ⟨*head*⟩ been" or "have ⟨*head*⟩ been"
adjective	"more ⟨*head*⟩" or "most ⟨*head*⟩" or 'very ⟨*head*⟩"

Figure 3: **Corpus pattern templates for part-of-speech clues.** The placeholders refer to word-forms derived from the headword: ⟨*plural*⟩ and ⟨*singular*⟩ are derived via morphology; ⟨*head*⟩ uses the headword as is.

Often the choice of speech parts for lexicalizations reflects idiosyncratic usages rather than just underlying semantics. To account for this, a set of features is included that is based on the relative frequency that the denotational headword occurs in contexts that are indicative of each of the main speech parts: singular, plural, count, mass, verbal, adjectival, and adverbial. See Figure 3. These patterns were determined by analyzing part-of-speech tagged text and seeing which function words co-occur predominantly in the immediate context for words of the given grammatical category. Note that high frequency function words such as 'to' were not considered because they are usually not indexed for information retrieval.

These features are derived as follows. Given a lexical assertion (e.g., (denotation HoundTheWord CountNoun 0 Dog)), the headword is extracted and then the plural or singular variant wordform is derived for use in the pattern templates. Corpus checks are done for each, producing a vector of frequency counts (e.g., ⟨29, 17, 0, 0, 0, 0, 0⟩). These counts are then normalized and then used as numeric features for the machine learning algorithm. Table 3 shows the results for the hound example and with a few other cases.

3.3 Sample criteria

We use decision trees for this classification. Part of the motivation is that the result is readily interpretable and can be incorporated directly by knowledge-based applications. Decision trees are induced in a process that recursively splits the training examples based on the feature that parti-

Head	Sing	Plural	Count	Mass	Verb	Adv	Adj
hound	.630	.370	0	0	0	0	0
book	.613	.371	.011	.001	0	.002	.001
wood	.577	.418	0	.004	0	.001	.001
leave	.753	.215	0	0	.024	.008	0
fast	.924	.003	0	.003	.001	.043	.027
stormy	.981	0	0	0	0	0	.019

Table 3: **Sample relative frequency values from corpus checks.**

```
if (genls Event) and
        (genls not ∈ {ConsumingFoodOrDrink,
            SeasonOfYear, QualitativeTimeOfDay,
            SocialGathering, PrecipitationProcess,
            SimpleRepairing, ConflictEvent,
            SomethingAppearingSomewhere}) and
        (isa not PhysiologicalConditionType) and
        (f-Plural ≤ 0.245) then
    if (Suffix ∈ {ine, een} then Verb
    if (Suffix ∈ {ile, ent} then CountNoun
    if (Suffix = ing) then MassNoun
    if (Suffix = ion) then
        if (f-Mass > 0.026) then MassNoun
        else Verb
    if (Suffix = ite) then CountNoun
    if (Suffix ∈ {ide, ure, ous} then Verb
    if (Suffix = ive) and
        (genls Perceiving) then MassNoun
    else CountNoun
    if (Suffix = ate) then
        if (not genls InformationStore) and
            (f-Count ≤ 0.048) and
            (f-Adverb ≤ 0.05) then
            if (gens Translocation) then MassNoun
            else CountNoun
```

Figure 4: **Sample rule from the general speech part classifier.**

tions the current set of examples to maximize the information gain (Witten and Frank, 1999). This is commonly done by selecting the feature that minimizes the entropy of the distribution (i.e., yields least uniform distribution). A fragment of the decision tree is shown to give an idea of the criteria being considered in the speech part classification. See Figure 4. In this example, the semantic types mostly provide exceptions to associations inferred from the suffixes, with corpus clues used occasionally for differentiation.

4 Evaluation and results

To test out the performance of the speech part classification, 10-fold cross validation is applied to each configuration that was considered. Except as noted below, all the results are produced using Weka's J4.8 classifier (Witten and Frank, 1999), which is an implementation of Quillian's C4.5 (Quinlan, 1993) decision tree learner. Other classifiers were considered as well (e.g., Naive Bayes and nearest neighbor), but J4.8 generally gave the best overall results.

4.1 Results for mass-count distinction

Table 4 shows the results for the special case mass-count classification. This shows that the system achieves an accuracy of 93.0%, an improvement of 24.4 percentage points over the standard baseline of always selecting the most frequent case (i.e., count noun). Other baselines are included for comparison purposes. For example, using the headword as the sole feature (*just-headwords*) performs fairly well compared to the system based on Cyc; but, this classifier would lack generalizability, relying simply upon table lookup. (In this case, the decision tree induction process ran into memory constraints, so a Naive Bayes classifier was used instead.) In addition, a system only based on the suffixes (*just-suffixes*) performs marginally better than always selecting the most common case. Thus, morphology alone would not be adequate for this task. The OpenCyc version of the classifier also performs well. This illustrates that sufficient data is already available in OpenCyc to allow for good approximations for such classifications. Note that for the mass-count experiments and for the experiments discussed later, the combined system over full Cyc leads to statistically significant improvements compared to the other cases.

4.2 Results for general speech part classification

Running the same classifier setup over all speech parts produces the results shown in Table 5. The overall result is not as high, but there is a similar improvement over the baselines. Relying solely on suffixes or on corpus checks performs slightly better than the baseline. Using headwords performs well, but again that amounts to table lookup. In terms of absolute accuracy it might seem that the system based on OpenCyc is doing nearly as well as the system based on full Cyc. This is somewhat misleading, since the distribution of parts of speech is simpler in OpenCyc, as shown by the lower entropy value (Jurafsky and Martin, 2000).

5 Related work

There has not been much work in the automatic determination of the preferred lexicalization part of speech, outside of work related to part-of-speech tagging (Brill, 1995), which concentrates on the

Dataset Characteristics

	OpenCyc	Cyc
Instances	2607	30676
Classes	2	2
Entropy	0.76	0.90

Accuracy Figures

	OpenCyc	Cyc
Baseline	78.3	68.6
Just-headwords	87.5	89.3
Just-suffixes	78.3	71.9
Just-corpus	78.2	68.6
Just-terms	87.4	90.5
Combination	88.4	93.0

Table 4: **Mass-count classification over Cyc lexical mappings.** *Instances* is size of the training data. *Classes* is the number of choices. *Entropy* characterizes distribution uniformity. *Baseline* uses more frequent case. The *just-X* entries incorporate a single type: *headwords* from lexical mapping, *suffixes* of headword, *corpus* co-occurrence of part-of-speech indicators; and Cyc reference *terms*. *Combination* uses all features except for the headwords. For Cyc, it yields a statistically significant improvement over the others at $p < .01$ using a paired t-test.

Dataset Characteristics

	OpenCyc	Cyc
Instances	3721	44499
Classes	16	34
Entropy	1.95	2.11

Accuracy Figures

	OpenCyc	Cyc
Baseline	54.9	48.6
Just-headwords	61.6	73.8
Just-suffixes	55.6	53.0
Just-corpus	63.1	49.0
Just-terms	68.2	71.3
Combination	73.1	77.8

Table 5: **Full speech part classification over Cyc lexical mappings.** All speech parts in Cyc are used. See Table 4 for legend.

sequences of speech tags rather than the default tags. Brill uses an error-driven transformation-based learning approach that learns lists for transforming the initial tags assigned to the sentence. Unknown words are handled basically via rules that change the default assignment to another based on the suffixes of the unknown word. Pedersen and Chen (1995) discuss an approach to inferring the grammatical categories of unknown words using constraint solving over the properties of the known words. Toole (2000) applies decision trees to a similar problem, distinguishing common nouns, pronouns, and various types of names, using a framework analogous to that commonly applied in named-entity recognition.

In work closer to ours, Woods (2000) describes an approach to this problem using manually constructed rules incorporating syntactic, morphological, and semantic tests (via an ontology). For example, patterns targeting specific stems are applied provided that the root meets certain semantic constraints. There has been clustering-based work in part-of-speech induction, but these tend to target idiosyncratic classes, such as capitalized words and words ending in '-ed' (Clark, 2003).

The special case of classifying the mass-count distinction has received some attention. Bond and Vatikiotis-Bateson (2002) infer five types of countability distinctions using NT&T's Japanese to English transfer dictionary, including the categories strongly countable, weakly countable, and plural only. The countability assigned to a particular semantic category is based on the most common case associated with the English words mapping into the category. Our earlier work (O'Hara et al., 2003) just used semantic features as well but accounted for inheritance of types, achieving 89.5% with a baseline of 68.2%. Schwartz (2002) uses the five NT&T countability distinctions when tagging word occurrences in a corpus (i.e., word tokens), based primarily on clues provided by determiners. Results are given in terms of agreement rather than accuracy; compared to NT&T's dictionary there is about 90% agreement for the fully or strong countable types and about 40% agreement for the weakly countable or uncountable types, with half of the tokens left untagged for countability. Baldwin and Bond (2003) apply sophisticated preprocessing to derive a variety of countability clues, such as grammatical number of modifiers, co-occurrence of specific types of determiners and pronouns, and specific types of prepositions. They achieve 94.6% accuracy using four categories of countability, including two categories for types of plural-only nouns.

Since multiple assignments are allowed, negative agreement is considered as well as positive. When restricted to just count versus mass nouns, the accuracy is 89.9% (personal communication). Note that, as with Schwartz, the task is different from ours and that of Bond and Vatikiotis-Bateson: we assign countability to word/concept pairs instead of just to words.

6 Conclusion and future work

This paper shows that an accurate decision procedure (93.0%) accounting for the mass-count distinction can be induced from the lexical mappings in the Cyc KB. The full speech part classifier produces promising results (77.8%), considering that it is a much harder task, with over 30 categories to choose from. The features incorporate semantic information, in particular Cyc's ontological types, in addition to syntactic information (e.g., headword morphology).

Future work will investigate how the classifiers can be generalized for classifying word usages in context, rather than isolated words. This could complement existing part-of-speech taggers by allowing for more detailed tag types, such as for count and agentive nouns.

A separate area for future work will be to apply the techniques to other languages. For example, minimal changes to the classifier setup would be required to handle Romance languages, such as Italian. The version of the classifier that just uses Cyc reference terms could be applied as is, given lexical mappings for the language. For the combined-feature classifier, we would just need to change the list of suffixes and the part-of-speech pattern templates (from Figure 3).

Acknowledgements

The lexicon work at Cycorp has been supported in part by grants from NIST, DARPA (e.g., RKF), and ARDA (e.g., AQUAINT). At NMSU, the work was facilitated by a GAANN fellowship from the Department of Education and utilized computing resources made possible through MII Grants EIA-9810732 and EIA-0220590.

References

Timothy Baldwin and Francis Bond. 2003. Learning the countability of English nouns from corpus data. In *Proc. ACL-03*.

Francis Bond and Caitlin Vatikiotis-Bateson. 2002. Using an ontology to determine English countability. In *Proc. COLING-2002*, pages 99–105. Taipei.

Eric Brill. 1995. Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics*, 21(4):543–565.

Kathy J. Burns and Anthony B. Davis. 1999. Building and maintaining a semantically adequate lexicon using Cyc. In Evelyn Viegas, editor, *Breadth and Depth of Semantic Lexicons*, pages 121–143. Kluwer, Dordrecht.

Alexander Clark. 2003. Combining distributional and morphological information for part of speech induction. In *Proceedings of EACL 2003*.

Daniel Jurafsky and James H. Martin. 2000. *Speech and Language Processing*. Prentice Hall, Upper Saddle River, New Jersey.

D. B. Lenat. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11).

Tom O'Hara, Nancy Salay, Michael Witbrock, Dave Schneider, Bjoern Aldag, Stefano Bertolo, Kathy Panton, Fritz Lehmann, Matt Smith, David Baxter, Jon Curtis, and Peter Wagner. 2003. Inducing criteria for mass noun lexical mappings using the Cyc KB, and its extension to WordNet. In *Proc. Fifth International Workshop on Computational Semantics (IWCS-5)*.

B. Onyshkevych and S. Nirenburg. 1995. A lexicon for knowledge-based MT. *Machine Translation*, 10(2):5–57.

Ted Pedersen and Weidong Chen. 1995. Lexical acquisition via constraint solving. In *Proc. AAAI 1995 Spring Symposium Series*.

Paul Procter, editor. 1995. *Cambridge International Dictionary of English*. Cambridge University Press, Cambridge.

J. Ross Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, California.

Lane O.B. Schwartz. 2002. Corpus-based acquisition of head noun countability features. Master's thesis, Cambridge University, Cambridge, UK.

Janine Toole. 2000. Categorizing unknown words: Using decision trees to identify names and misspellings. In *Proc. ANLP-2000*.

Ian H. Witten and Eibe Frank. 1999. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, CA.

W. Woods. 2000. Aggressive morphology for robust lexical coverage. In *Proc. ANLP-00*.